

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 15:50:54 ; Search time 2492.14 Seconds
(without alignments)
16465.781 Million cell updates/sec

Title: US-09-497-967-102

Perfect score: 1410

Sequence: 1 atgaagaacacatctcgtg.....cttactacctgctgtaataa 1410

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_man.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	784.4	55.6	3026	3	AF324424
2	63.2	4.5	2486	3	AF140273
3	57.2	4.1	1520	3	AF405431
4	55.2	3.9	125020	9	AF429315
5	52.2	3.7	1249	9	ICVIMANT
6	50.0	3.5	125020	9	AF429315
c	43.4	3.1	230372	2	AC073693
7	41.4	2.9	159764	9	AC093866
8	41.0	2.9	643	8	AF374493
9	41.0	2.9	179217	2	AC099748
10	39.4	2.8	591	8	AF374552
11	38.6	2.7	643	8	AF374492
12	38.6	2.7	16814	9	HSCOLPDGF
c	38.6	2.7	341950	1	AP003600
13	38.4	2.7	224271	9	AC026803
c	38.2	2.7	15294	1	AE006936
14	38.2	2.7	31300	1	MTV035
15	38.2	2.7	152777	2	AC125855
c	38.2	2.7	233345	4	AJ421481
16	38.2	2.7	338150	1	AP005278
17	38.2	2.7	349980	6	AX127147
18	37.8	2.7	559	8	AF374543
19	37.8	2.7	591	8	AF374540
20	37.8	2.7	591	8	AF374542
21	37.8	2.7	591	8	AF374544
22	37.8	2.7	591	8	AF374545
23	37.8	2.7	591	8	AF374547
24	37.8	2.7	591	8	AF374548
25	37.8	2.7	591	8	AF374549
26	37.8	2.7	591	8	AF374550
27	37.8	2.7	591	8	AF374551
28	37.8	2.7	643	8	AF374490
29	37.8	2.7	643	8	AF374491
30	37.8	2.7	643	8	AF374494
31	37.8	2.7	643	8	AF374495
32	37.8	2.7	643	8	AF374496
33	37.8	2.7	643	8	AF374497
34	37.8	2.7	643	8	AF374498
35	37.8	2.7	643	8	AF374499
36	37.8	2.7	643	8	AF374500
37	37.8	2.7	643	8	AF374501
38	37.8	2.7	643	8	AF374502
39	37.8	2.7	643	8	AF374503
40	37.8	2.7	643	8	AF374504
41	37.8	2.7	643	8	AF374505

ALIGNMENTS

RESULT 1
AF324424
LOCUS
DEFINITION
ICthyophthirius multifiliis immobilization antigen isoform
(IAG52A) gene, complete cds.
ACCESSION
AF324424
VERSION
AF324424.1
KEYWORDS
GI:12698726
SOURCE
ICthyophthirius multifiliis.
ORGANISM
ICthyophthirius multifiliis.
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomatida; Ophryoglenina; Ichthyophthirius.
REFERENCE
1 (bases 1 to 3026)
AUTHORS
Lin, Y., Lin, T.L., Wang, C.C., Wang, X., Stieger, K., Klopfleisch, R.
and Clark, T.G.

Pred. No. is the number of results predicted by chance to have a

TITLE		Variation in primary sequence and tandem repeat copy number among	
JOURNAL	i-antigens of Ichthyophthirius multifiliis		
MEDLINE	Mol. Biochem. Parasitol. 120 (1), 93-106 (2002)		
PUBMED	21839613		
REFERENCE	11849709		
AUTHORS	2 (bases 1 to 3026)		
TITLE	Lin, Y., Lin, T.-L. and Clark, T.G.		
JOURNAL	Direct Submission		
	Submitted (27-NOV-2000) Microbiology and Immunology, NYSCVM Cornell University, Ithaca, NY 14853, USA		
FEATURES	Location/Qualifiers		
source	1..3026		
	/organism="Ichthyophthirius multifiliis"		
	/macronuclear		
	/serotype="D"		
	/isolate="G5"		
	/db_xref="taxon:5932"		
gene	1285..2817		
	/gene="IAG52A"		
mRNA	1285..2817		
	/gene="IAG52A"		
	/product="immobilization antigen isoform"		
CDS	1330..2736		
	/gene="IAG52A"		
	/note="52 kDa protein"		
	/codon_start=1		
	/transl_table=6		
	/product="immobilization antigen isoform"		
	/protein_id="AAK01661.1"		
	/db_xref="GI:12698727"		
	/translation="MKNNILIVILISLFINQIKSAPCVGTETAGQVDDLGTPANC		
	VNCKNFYNNAAAFVPGASTCTPCPKKDAGAPNPATANLVTCNVKPAATIAIA		
	GGATDVAIITECVNCRINPYENADNFNAGASTCTACPNRVGGALTAGNAATIAVG		
	CNVACPTGALDGTITDYVRSCTECVKCRNLNYYNGNNGTFFNPKGKSCTCPCLPK		
	PANVACATLGNDAITTCAGACPDGTISAAGVNVMQAQTECTNAPNFINNAPNF		
	NPQNSTCLCPANKDYGAEATAGGAATLAKQCNLACPDGTIAISGATNYVILQTECLN		
	CAANFYPDNGNFQSGSRKACPAKVGAVATAGGTATLIAQCALECPAGTVLTGDT		
	TSYKQAAASBCVKCAANFYTKOTDWMAGIDTCTSCNKLKLTSGAEANLPBSAKKNQIQ		
	DFANFISLSLLLSYLL"		
BASE COUNT	1177 a 316 c 355 g 1178 t		
ORIGIN			
Query Match	55.6%	Score 784.4;	DB 3; Length 3026;
Best Local Similarity	72.3%	Pred. No. 3.9e-220;	
Matches 1019; Conservative	0;	Mismatches 391;	Indels 0; Gaps 0;
Qy 1	ATGAAGAACAACATCTGGTGATCCTGATCATCTCTCTGTTTCATCAACCATCAAGTCT	60	
Db 1330	ATGAAAATAATATTTTAGTAATATTGATTTATTTTATTTATTTATTTATTTATTTATTT	1389	
Qy 61	GCTAACTCTCTGTGGGACCGAGACCAACACGCTGGACAGGTGGACCTGGGAACC	120	
Db 1390	GCTAATTGCTCTGTGAACTTGAACCTTACACACGCGGTAAGTTGATGATCTAGGAAT	1449	
Qy 121	CCTGCTAACTGTGAACCTGTGAGAAGAACTTCTACTACAAACGCTGCTGCTTTTG	180	
Db 1450	CCTGAAATTTGTTAATTTGTAGAAAACTTTTATTAATAATGCTGCTGCTTCGTT	1509	
Qy 181	CCTGGAGCTTCTACTGTACCCCTTGCTCTCAGAAAGAGACGCTGGAGCTCAGCCTAAC	240	
Db 1510	CCTGCTGTAGTACGCTGTACACCTTTGTCATAAAAAAGATGCTGGTGTCTTAACCAAT	1569	
Qy 241	CCTCCTGTACCGCTAACCTGGTGACCGAGTAACTGAGTGTCTGCTGGGAACCGCT	300	
Db 1570	CCACTGCTACTGCTAATTTAGTCACATAATGTAACTGTTAATGCCCTGTGGTACCGCA	1629	
Qy 301	ATCGCTGGAGGAGCTACCGACTAGCGTGCTATCATCACCGAGTGTGTGAACGTGCGCATC	360	
Db 1630	ATTGAGGTGGCAACAGATTATGCAGCATAATCACAGAATGTGTTAATTGTAGAATT	1689	
Qy 361	AACCTTCTACAACGAGACGGCTCCTAACTTCAACGGTGGAGCTTCTACCTGACCGCTTGT	420	
Db 1690	AATTTTTATAATGAAATGCTCCAAATTTTAAATGCAGGTGCTAGTACATGACAGCTTGT	1749	

QY	421	CCTGTGAACCGCTGGAGAGACTCTGACCGCTGGAACGCTGCTACCATGCTGGCTCAG	480
DB	1750	CCGGTAAACAGATTGGTGGTCATTGACTGCTGTGTAATGGCGTTACCATAGTCGCATAAA	1809
QY	481	TGTAACGTGCTTGTCTACCGGAACGCTCTGGAGCAGCGAGTGAACACGACTACGCTG	540
DB	1810	TGTAACGTGCGATGTCTACTGTGTACTGCATTGATGATGGAGTAACCTACTGATTATGTT	1869
QY	541	CGCTCTTTTCCACCGAGTGTGAAGTGTCCCTGGAACCTTCTACTACAAGGAAACACGGA	600
DB	1870	AGATCATTCACAGAAATGTGTTAAATCTAGACTTAACCTTACTATAATAGTAAATGGT	1929
QY	601	AACACCCCTTTCAACCCCTGGAAGTCTAGTGTACCCCTGTGCTGCTATCAACGCCCTGCT	660
DB	1930	AATACTCCTTTCAATCCAGGTAAAGTTAATGCACACCTTGTGCGGCAATTTAAACACCT	1989
QY	661	AACGTGGCTCAGACTACCCCTGGGAACAGCTACCATCACCGCTACGTGTGAACGTGGCT	720
DB	1990	AATGTGCTTAAGCTACTTTAGGTAATGATGCTACAAACCGCAATATGTAAAGCTTGCA	2049
QY	721	TGTCCTTGACGGAACATCTCTGCTGTGGAGTGAACAACCTGGGTGGCTCAGAACCCGAG	780
DB	2050	TGCCCTGATGTAATAAGTGTCTGCGAGTAATAATTTGGTAGCAAAACACACGTAA	2109
QY	781	TGTACCAACTGTGCTCTTAACCTCTACACACACAGCTCTCTAAGCTTCAACCCCTGGAAC	840
DB	2110	TGTACTAATTTGTGCTCTAAGCTTTTACAATAATAATGCTCTTAATTTCAATPCAGAGTAAT	2169
QY	841	TCTACCTGTCTGCTTGTCTGCTAACAAGGACTACGGAGCTGAGGCTACCGCTGGAGGA	900
DB	2170	AGTACATGCGCTACCTTGCCGAGCAATAAAGATTTATGGTGTGAAGCCACTGCGAGTGGT	2229
QY	901	GCTGCTACCCCTGGCTAAGCAGTGTAAACATCGCTTGTCTGACGGAACCGCTATCGCTTCT	960
DB	2230	GCCGCTACTTTAGCCAATAATGTAATAATTGCATGCCCTGATGGTACTGCAATTTGCTAGT	2289
QY	961	GGAGCTACCAACTACGTGATCTCTGAGACCGGAGTGTCTGAACTGTGCTTAACCTTCTAC	1020
DB	2290	GGAGCACTAATTTATGTAATTTATAAACAGAAATGTCTAAATTTGTGCTTAACCTTTTAT	2349
QY	1021	TTTCAGCGAAACAACTTCCAGGCTGGATCTTCTCGCTGTAAAGGCTGTCTGCTTAACAAG	1080
DB	2350	TTTGAATGGTAATTAATTTCTAGGAGGAGTAAGTACTAGATGCAAGCATGTCCAGCAATAAA	2409
QY	1081	GTGCAGGAGCTGTGGCTACCGCTGGAGAACCGGTACCTGTATCGCTCACTGTGCTCTG	1140
DB	2410	GTTTAAAGCGCTGTAGCAACTGCAGGTGTACTGTCTACTTTAATTTGCAATAATGTGCCCTT	2469
QY	1141	CAGTGTCTCTGGAACCGGTGTACCGGAGGAACCACTCTACCTACAAGCAGGCTGCT	1200
DB	2470	GAATGCCCTGCTGTACTGTACTACCGATGGAAACAACATCTACTTATAATAGCAGCA	2529
QY	1201	TCTGAGTGTGTGAAGTGTGCTTAACCTTCTACACCAACAGCAGACCGCTGGGTGGCT	1260
DB	2530	TCTGAATGTGTTAAATGTGCTGCCAAGCTTTTACTACAAAATAAAGCTGATTTGGGTAGCA	2589
QY	1261	GGATCGACACCTGTACCTCTTGTACACAGAAGCTGACCTCTGGAGCTGAGGCTAACCTG	1320
DB	2590	GGTATTGATACATGTACTAGTTGCTTAATAAAAAAATAAAGCTTCTGGCGCTGAAGCTAATTTA	2649
QY	1321	CCTGAGTGTGTTAAGAGAACAATCCAGTGTGACTTTCGGTAACTTCTGCTATCTCTCTG	1380
DB	2650	CCTGAATCTGCTAAAAAATAATATAATGTGATTTCCGTAAATTTTTTATCAATTTCCCTTA	2709
QY	1381	CTGCTGATCTCTACTACCTGCTGTAATAA	1410
DB	2710	TTAATGATTTCTTAATTAATTTATATCATGA	2739

RESULT 2
AF140273
LOCUS

DEFINITION Ichthyophthirius multifiliis immobilization antigen precursor (IAG48) gene, complete cds.

ACCESSION AF140273

VERSION AF140273.1 GI:4868370

KEYWORDS

SOURCE Ichthyophthirius multifiliis.

ORGANISM Ichthyophthirius multifiliis

REFERENCE 1 (bases 1 to 2486) Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida; Ophryoglenina; Ichthyophthirius.

AUTHORS Clark, T.G., Lin, T.-L., Jackwood, D.A., Sherrill, J., Lin, Y., and Dickerson, H.W.

TITLE The gene for an abundant parasite coat protein predicts tandemly repetitive metal binding domains

JOURNAL Gene 229 (1-2), 91-100 (1999)

MEDLINE 99196987

PUBMED 10095108

REFERENCE 2 (bases 1 to 2486) Gaertig, J., Gao, Y., Tishgarten, T., Clark, T.G., and Dickerson, H.W. Surface display of a parasite antigen in the ciliate Tetrahymena thermophila

JOURNAL Nat. Biotechnol. (1999) In press

REFERENCE 3 (bases 1 to 2486) Clark, T.G., Lin, T.-L., Jackwood, D.A., and Dickerson, H.W. Direct Submission

AUTHORS

TITLE Submitted (29-MAR-1999) Microbiology & Immunology, Cornell University, College of Veterinary Medicine, Ithaca, NY 14853, USA

JOURNAL

FEATURES

source

1..2486

/organism="Ichthyophthirius multifiliis"

/serotype="A"

/isolate="GI"

/db_xref="taxon:5932"

399..1845

/gene="IAG48"

399..1845

/gene="IAG48"

/product="immobilization antigen precursor"

433..1761

/note="IAG48"

/note="surface protein; 48 kDa i-antigen"

/codon_start=1

/transl_table=6

/product="immobilization antigen precursor"

/protein_id="AAD31283.1"

/db_xref="GI:4868371"

/translation="MKYNLLILILISLFINELRAVPCPDGTQTQAGLTDVGAADLGTC VNCRPNEYNGAAGANGQPAANNAARGICVPCQINRVGVTNAGDLATLQTC STQCTGTALDDGVTDFDRSAAQCCKPNFYNGSGPQGEAFGVQVFAAGAAAGV AAVTSQCVPCLNKNDSPATGAQAANLQCSNQCPTGLDDGVTLVFNTSATLCVK CRPNFYNGSGPQGEAFGVQVFAAGAAAGVAAVTSQCVPCLNKNDSPATGAQAANL ATQCSNQCPTGTATQDQVTLVFSNSQCSOCIANFYFNFGNFEAGKSOCLKCPVSKTT PAHAPGNTATQATCLTCTPAGTVLDGTSNFEVASATECTKSAGFSBKTFTAG TDTCTCKKLTSATAKVYAEATQKVCQASTTFPAKFLISLISFISFYLL"

433..492

/gene="IAG48"

493..1758

/gene="IAG48"

/product="immobilization antigen"

BASE COUNT 896 a 310 c 321 g 959 t

ORIGIN

Query Match 4.5%; Score 63.2; DB 3; Length 2486;

Best Local Similarity 52.2%; Pred. No. 4.5e-07;

Matches 140; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 1098 TACCGCTGAGAACCGTACCCCTAGTCGCTCAGTGTGCTCTGGAGTGTCTCTGCTGGAAC 1157

Db 1443 TGCTCCAGGTAATCTGCTACTTAAGCCACATAATGTTTGACCAACATGCTCTGCTGTAC 1502

QY 1158 CTTGTGACCGACGGAACCACTTACCTACACAGCAGCTCTTCTGAGTGTGGAAGT 1217

Db 1503 AGTACTTGATGATGACATCACTAATTTGTAGTTCGCGCACTGAATGCTAAATG 1562

QY 1218 TGCTGCTAACTTCTACACCACCAAGCAGCAGCAGCTGGTGGCTGAATCGACACCTGTAC 1277

Db 1563 TTCTGCTGGCTTTTGGCATCAAAACAACTGGTTTACAGCAGGTACTGATCATGTAC 1622

QY 1278 CTTCTGTACAGAAGCTGACCTCTGGAGCTGAGCTAACCTGCTGAGTGTGCTGAAGA 1337

Db 1623 TGAATGTACTAAAAAATAACTTCTGGTCCACAGCTAAAGTATATGCTGAAGCTACTCA 1682

QY 1338 GAACATCCAGTGTGACTTCGCTAACTTC 1365

Db 1683 AAAAGTATAATGCGCTCCACTACTTTC 1710

RESULT 3

AF405431

LOCUS

DEFINITION Ichthyophthirius multifiliis 1520 bp mRNA linear INV 27-FEB-2002

protein mRNA, complete cds.

ACCESSION AF405431

VERSION AF405431.1 GI:15290741

KEYWORDS

SOURCE Ichthyophthirius multifiliis.

ORGANISM Ichthyophthirius multifiliis

REFERENCE 1 (bases 1 to 1520) Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida; Ophryoglenina; Ichthyophthirius.

AUTHORS Lin, Y., Lin, T.-L., Wang, C.-C., Wang, X., Stieger, K., Klopfleisch, R., and Clark, T.G.

TITLE Variation in primary sequence and tandem repeat copy number among i-antigens of Ichthyophthirius multifiliis

JOURNAL Mol. Biochem. Parasitol. 120 (1), 93-106 (2002)

MEDLINE 21839613

PUBMED 11849709

REFERENCE 2 (bases 1 to 1520) Lin, Y., Wang, J.-C., and Clark, T.C. Direct Submission

AUTHORS

TITLE Submitted (03-AUG-2001) Microbiology and Immunology, NYSCVM Cornell University, Ithaca, NY 14853, USA

JOURNAL

FEATURES

source

1..1520

/organism="Ichthyophthirius multifiliis"

/macronuclear

/serotype="D"

/isolate="G5"

/db_xref="taxon:5932"

41..1423

/note="IAG52B[G5]"

/codon_start=1

/transl_table=6

/product="52kDa immobilization antigen variant B protein"

/protein_id="AAK94941.1"

/db_xref="GI:15290742"

/translation="MKFNILILILISLFINELRAVNCNPAANGOSDPTGAADINTC THCKHFYNGGNPAGAPGVQVNSQICACQVHKADSHRGGDANLAQSNL CPAGTAVEDSGSPFTFQSLQCVNCPNFYNGNPNATQCNNECPTGTATQDGAIFTYQ NPEVPNVSSPGQCVACQVKNQSDSLRPAQANLATQCNNECPTGTATQDGAIFTYQ SISQCTECKVDVDFYNGNPNQNGQFTPGQLIANPDAAATAGQIPVPGNSKQVA CESKKTNSRSGLANLAQCGTECPAGTLVDGVTPTTVLSQCVNPKAGFYONS NFEAGKSCNKCASVSKTGSASVPGNSATSATOCNDPCAGTVDDGTSTNRFVALSEC TKQANFYASKTSGFAAGTDTCTCECKSKLTGATKAVYAEATQKACASSTPAKFLSM SLIFISFYLL"

493 a 255 c 256 g 516 t

BASE COUNT

ORIGIN

Query Match 4.1%; Score 57.2; DB 3; Length 1520;

Best Local Similarity 52.0%; Pred. No. 2.5e-05;

Matches 159; Conservative 0; Mismatches 138; Indels 9; Gaps 1;

QY 1114 GCTACCTGATCGCTCAGTGTGCTGTGGAGTGTCTCTGCTGAACCGCTGCTGACCGACGA 1173

Db 1121 GCTACTTCAGCCACATAATGTTAAACGATGTCCTGCTGGTACAGTGGTGTGATGATGCT 1180

```
QY 1174 ACCACCTCTACCTACCAAGCAGCTCTCTCTGAGTGTGTGAAGTGTGCTGCTAACTTCTAC 1233
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1181 ACATCAACTAATTTTGTAGCTTTAGCAAGTGAATGTACTAAATGTTAGGCTAACTTTAT 1240

QY 1234 ACCACCAAGCAGCAGCTGGTGGCTGGAATCGACACCTGTACCTCTGTGTAACAAGAA 1293
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1241 GCATCAAAACATCTGGTTTTCGACAGGACTGTATACATGTACATGATGTCTTAAAAAA 1300

QY 1294 CTGACCTCTGGAGCTGAGGCTAACTGCTGAGTGTCTGTAAGAAACATCCAGTGTGAC 1353
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1301 TTAACCTCTGGTGTACACCTAAAGTATATCTGGAAGCTACTTAAAAAGCAATATGCGCC 1360

QY 1354 -----TTGCGTAACCTCTGCTATCTCTGCTGCTGCTGATCTCTACTACCTGCTG 1404
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1361 AGTTCACCTTTCGCAAAATTTTATCAATGCTCTTAATATTATTTCTTCTATTGTTG 1420

QY 1405 TAATAA 1410
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1421 TGATGA 1426

RESULT 4
AF429315 125020 bp DNA linear PRI 18-JAN-2002
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 125020)
Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
Nat. Genet. 29 (4), 377-378 (2001)
21583737
11694876
2 (bases 1 to 125020)
Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
Direct Submission
Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
Location/Qualifiers
    1..125020
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="16"
    /map="16q24.3; between D16S520 and WI-12410"
    /note="isolated from a patient with Huntington's
    Disease-like 2 (HDL2)"
    complement(35581..35746)
    /rpt_type=tandem
    /rpt_unit=ctg
    /complement(<36507..>36887)
    /gene="JPH3"
    /note="JPH3"
    complement(<36507..>36887)
    /gene="JPH3"
    /product="junctophilin 3"
    complement(<36507..36887)
    /gene="JPH3"
    /note="component of the junctional complex between plasma
    membrane and endoplasmic reticulum"
    /codon_start=1
    /product="junctophilin 3"
    /protein_id="AAL40941.1"
    /db_xref="GI:17646245"
    /translation="MSSGGRFNFDGGSYCGGWEDGKAHGHGVCYTPKGGQGYTGSWS
    HGFVLGVYTPSGNTYQGTWAGQKRHGIGLESKQKVVYKGEWTHGFKRGYGVRECA
    G"

repeat_region
gene
mRNA
CDS
```

```
BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others
ORIGIN
Query Match 3.9%; Score 55.2; DB 9; Length 125020;
Best Local Similarity 10.9%; Pred. No. 0.00014;
Matches 108; Conservative 421; Mismatches 452; Indels 6; Gaps 3;
NGAKYEGTWSNGLQDGYGTETYSDG"
QY 199 ACCCTTGTCTCAGAGAAGGACGCTGGAGCTCAGCTAACCTCTGCTACCGCTAAC 258
    || || || || || || || || || || || || || || || || || || || || || ||
Db 16868 ATCCARKKMKSHAGRRRCYYSKSWMBMSVSVSKMHSHASBSCHWBKCMWSC 16927

QY 259 CTGCTGACCCAGTGAAGTGAAGTCTCTGCTGGAACCGCTATCGCTGGAGGAGTACC 318
    || || || || || || || || || || || || || || || || || || || || || ||
Db 16928 CMSMMYKSSWGWSSWGCWRRRSKWKWYVRGMSRSKSMRYTGGSKMMRSMCTSSC 16987

QY 319 GACTACGCTGTATCATCACCAGGAGTGTGAACCTGTCGCATCAACTTCTACAACGAGA 378
    || || || || || || || || || || || || || || || || || || || || || ||
Db 16988 YASMCMMCCWSCCMRSCCMRSYCCCMRYCCACKCYMSYWTMSASYMSRSRYWRS 17047

QY 379 GCTCTTAACCTCAACGCTGGAGCTTCTACCTGTACCGCTTGTCTGTGAACCGCGTGG 438
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17048 KCWS--RMCSSRSRCKRCRCGSGMGKGGKSGYGRKTRKRSKMGKAKWYMYRRSM 17105

QY 439 GGAGCTCTGACCGCT---GGAACGCTGCTACCATCTGCTGCTCAGTGTACGCTGGCTGT 495
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17106 RKMVSKSGMYCMYCWGRRRCYCSCTSRSAMCCSYCAKCKSMCYSCYGYMSMKGY 17165

QY 496 CTTACCGGAACCGCTCTGGACGAGGAGTACCACCGACTACGTGCGCTCTTTTCAACGAG 555
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17166 YYACSYRGSMSKSKYMRGTYSTSGGCCCTTTTCCCCCNANTGGGAAGCTTTTNCNK 17225

QY 556 TGTGTGAAGTGTGCTGAACTTCTACTACAAACGGAACAAACGGAACACCCCTTTTCAAC 615
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17226 TSYYRKRNGCAMCKYNNYNNWSRSCRAGNSCTTKYKSSMTSMASVCMWSMYCSMR 17285

QY 616 CTTGAAAGTCTCAGTGTACCCCTTGCTCTCTATCAAGCTGCTTAAGTGGCTCAGCT 675
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17286 SMASSGMSWSYMKMKSSWRMSYMKCCMKCYCMRMSRMSRMSRMSRMSRMSR 17345

QY 676 ACCCTGGGAAACGAGCTACCATCACCGCTCAGTGTAAACGCTGCTGTCTGACGGAAC 735
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17346 GCYCTRCYWMSSSKSKSYKSYMMMSKRMKMGKWSRCWGSWMSASRSCKYKYSMRCS 17405

QY 736 ATCTCTGCTGTGGAGTGAACAACACTGGGTGGCTCAGACACCGAGTGTACCACTGTGCT 795
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17406 MMSSRCYRCAGCMKMGYMYRMCWSMRKRRWKGKWSAMYRMRMMKMGKAMMKCYSR 17465

QY 796 CTTAACTTCTACAACAACGCTCTCTAACTTCAACC-CTGGAACACTCTACCTGTCTGCC 854
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17466 MSRCMMMKSKYCASRSCAMSGTYMKCASVCMSSYWCCTWCCTSYMSYRCTCCKGWS 17525

QY 855 TTGCTCTGCTAACAAGGACTACGAGCTAGGCTACCGCTGGAGGAGTGTCTACCTGGC 914
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17526 SVTKSKSSSSMSYKGRKRSYSCCTSRGAMSCWRRCYMRGASSMRGASSRRRAK 17585

QY 915 TAAGCAGTGTACATCGCTTGTCTGACGGAACCGCTACGCTCTGGAGCTACCAACTA 974
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17586 RSWGGRSKWMTGGMWRSKYIYCTGRRMMYIMCCRRRSYMYRMSAMGRKSSWS 17645

QY 975 CGTGATCTCTGACAGCGAGTGTCTGAACCTGTGCTCTACTTCTACTTCGACGGAACAA 1034
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17646 GMRMGSASSRRCKSASRSRSCWSSRRMKMRGSCWSSKMGWSRSASCKKSGRMR 17705

QY 1035 CTTCCAGGCTGATCTTCTCGCTGTAAAGCTTGTCTCTGCTAACAAGTGTACAGGAGTGT 1094
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17706 SKRSSYRKRGRGKRSMTKSKGSKGSKCWRKRSMTSSCYYSASSCMWMSKSCMC 17765

QY 1095 GCTACCGCTGGAGGAACCGCTACCTGATCGCTCAGTGTGCTGCTGAGTGTCTCTGCTGG 1154
    || || || || || || || || || || || || || || || || || || || || || ||
Db 17766 CMMMKRRACCCYSSNSCTSMYRCWGGMKSYSTGCTTCKCTGYKKSRTWYWMYSWTS 17825

QY 1155 AACCGTGTCTGACCGGAAACACCTC 1181
```



```

* 65521 55620: gap of unknown length
* 65621 80284: contig of 14664 bp in length
* 80284 80384: gap of unknown length
* 80384 88449: contig of 18065 bp in length
* 88449 98549: gap of unknown length
* 98549 116262: contig of 17713 bp in length
* 116262 116362: gap of unknown length
* 116362 140706: contig of 24344 bp in length
* 140706 140806: gap of unknown length
* 140806 167609: contig of 26803 bp in length
* 167610 167709: gap of unknown length
* 167710 199398: contig of 31689 bp in length
* 199399 199498: gap of unknown length
* 199499 230372: contig of 30874 bp in length.
FEATURES             Location/Qualifiers
     source           1. .230372
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-152L22"
BASE COUNT           61266 a 53740 c 52522 g 61042 t 1802 others
ORIGIN
Query Match          3 1%; Score 43.4; DB 2; Length 230372;
Best Local Similarity 53.9%; Pred. NO. 0.46;
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 166 GCTGCTGTTTGGTGGAGCTTCTACCTGTACCCCTTGTCTCAGAGAAGGAGCT 225
||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212511 GCTGCTGCTCCCTGGCTGCTGCTCTCCCTGGCTGCTGCTGCTGCTGCT 212452

QY 226 GGAGCTACCCCTAACCTCTGCTACCGCTAACCTGTGACCCAGTGTAAAGTGT 285
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212451 GCTGCTGCTCCCTGGCTGCTGCTACTGCTCCCTGGATGCTGCTCCCTGAATGGT 212392

QY 286 CTTGCTGGAAACCGCTATGCTGAGGAGCTACCGACTACGCTGCT 330
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212391 GCTACTGCTCCCTGGCTGCTGCTGCTGCTCCCTGGATGCTGCT 212347

RESULT 8
AC093866          159764 bp      DNA      linear      PRI 01-MAR-2002
LOCUS             Homo sapiens BAC clone RP11-582C12 from 4, complete sequence.
DEFINITION        AC093866 AC025608
ACCESSION          AC093866.3 GI:16973777
VERSION            HTG.
KEYWORDS
SOURCE             Homo sapiens.
ORGANISM           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE          1 (bases 1 to 159764)
AUTHORS            Sulston, J.E. and Waterston, R.
TITLE             Toward a complete human genome sequence
JOURNAL            Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE            99063792
PUBMED            9847074
REFERENCE          2 (bases 1 to 159764)
AUTHORS            Pearman, C., Meyer, R. and Doeber, A.
TITLE             The sequence of Homo sapiens BAC clone RP11-582C12
JOURNAL            Unpublished (2001)
REFERENCE          3 (bases 1 to 159764)
AUTHORS            Waterston, R.H.
TITLE             Direct Submission
JOURNAL            Submitted (10-SEP-2001) Genome Sequencing Center, Washington
MO 63108, USA
REFERENCE          4 (bases 1 to 159764)
AUTHORS            Waterston, R.H.
TITLE             Direct Submission
JOURNAL            Submitted (17-NOV-2001) Genome Sequencing Center, Washington
MO 63108, USA

```

repeat_region	646. 735	/rpt_family="ERV1"
repeat_region	811. 1176	/rpt_family="MaLR"
repeat_region	2191. 2508	/rpt_family="MaLR"
repeat_region	2504. 2654	/rpt_family="MER2_type"
repeat_region	2771. 3012	/rpt_family="MER2_type"
repeat_region	3865. 3900	/rpt_family="MaLR"
repeat_region	4553. 4848	/rpt_family="(CA)n"
repeat_region	5091. 5156	/rpt_family="Alu"
repeat_region	5387. 5515	/rpt_family="GA-rich"
repeat_region	5562. 5671	/rpt_family="Alu"
repeat_region	6206. 6570	/rpt_family="L2"
repeat_region	6680. 7334	/rpt_family="MaLR"
repeat_region	7335. 7712	/rpt_family="L1"
repeat_region	7713. 7960	/rpt_family="MaLR"
repeat_region	7961. 8248	/rpt_family="L1"
repeat_region	8249. 8674	/rpt_family="Alu"
repeat_region	8771. 8846	/rpt_family="L1"
repeat_region	8920. 8977	/rpt_family="MIR"
repeat_region	9169. 9212	/rpt_family="CT-rich"
repeat_region	9647. 9709	/rpt_family="AT-rich"
repeat_region	10764. 11375	/rpt_family="L2"
repeat_region	11377. 11733	/rpt_family="L1"
repeat_region	11741. 12592	/rpt_family="MaLR"
repeat_region	12593. 12815	/rpt_family="L1"
repeat_region	12816. 13318	/rpt_family="MER1_type"
repeat_region	13323. 13372	/rpt_family="L1"
repeat_region	13830. 14371	/rpt_family="MER1_type"
repeat_region	14621. 14680	/rpt_family="ERV1"
repeat_region	14694. 14816	/rpt_family="L1"
repeat_region	15382. 15437	/rpt_family="Alu"
repeat_region	15438. 15875	/rpt_family="MER1_type"
repeat_region	15876. 16041	/rpt_family="ERV1"
repeat_region	16406. 16850	/rpt_family="MER1_type"
repeat_region	17833. 18128	/rpt_family="ERV1"
repeat_region	18115. 18138	/rpt_family="MaLR"
repeat_region	18134. 18238	/rpt_family="AT-rich"
repeat_region	18217. 18239	/rpt_family="Alu"

repeat_region	/rpt_family="(TAAA)n"
repeat_region	/rpt_family="L1"
repeat_region	/rpt_family="L1"
repeat_region	/rpt_family="AT-rich"
repeat_region	/rpt_family="AT-rich"
repeat_region	/rpt_family="Achobo"
repeat_region	/rpt_family="L1"
repeat_region	/rpt_family="AT-rich"
repeat_region	/rpt_family="L1"
repeat_region	/rpt_family="AT-rich"
repeat_region	/rpt_family="AT-rich"
repeat_region	/rpt_family="Achobo"
repeat_region	/rpt_family="AT-rich"

Query Match

Best Local Similarity

Matches

Conservative

0;

Mismatches

81;

Indels

0;

Gaps

0;

QY

82

GAGACCAACACCGCTGGACAGGTGGACGACCTGGGAGCCCTGCTTAACCTGCTGCTGAACCTGT

141

Db

133672

GAAGCCTTCAGAGAGATGGGTGACTGCTGGGTGCCAGGATAGGGGTTCCACAGA

133731

QY

142

CAGAAGAACTTCTACTACAACAACGCTGCTGCTTGTGCTGGAGCTTCTACCTGTACC

201

Db

133732

GTAAGGACATCTACTGCTACAACCATGTTCTTGGAGCTATAAAGTGAAC

133791

QY

202

CGTTGCTCACAAGACGCTGGAGCTACGCTACCCCTCCCTGCTACC

252

Db

133792

TATAGTCATCAGCAGCTGAATATGAGGGTGAGAAACATCTATGGACC

133842

RESULT

9

AF374493

LOCUS

DEFINITION

AF374493

AF374493.1

GI:14336349

VERSION

KEYWORDS

SOURCE

ORGANISM

Thalassiosira weissflogii.

Thalassiosira weissflogii

Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;

Thalassiosirophyceae; Thalassiosirales; Thalassiosiraceae;

Thalassiosira.

1 (bases 1 to 643)

Armbrust,E.V. and Galindo,H.M.

Rapid evolution of a sexual reproduction gene in centric diatoms of

the genus Thalassiosira

Appl. Environ. Microbiol. 67 (8), 3501-3513 (2001)

JOURNAL

MEDLINE

PUBMED

11472926

2 (bases 1 to 643)

Armbrust,E.V. and Galindo,H.G.

Direct Submission

Submitted (27-APR-2001) Oceanography, University of Washington, Box

357940, Seattle, WA 98195-7940, USA

Location/Qualifiers

1. .643

/organism="Thalassiosira weissflogii"

/isolate="CCMP 1336"

/db_xref="taxon:67004"

/clone="10"

/country="USA: Long Island, NY"

mRNA

```
join(<1..56,141..>643)
/product="sexually induced protein SIG 1"
join(<1..56,141..>643)
/codon_start=2
```

CDS

```
/product="sexually induced protein SIG 1"
/protein_id="AAK61135.1"
/db_xref="GI:14336350"
/translation="SHDMCTCDNRNFGAGCSLRTPFGKAHVDTPRGDLASLTIGDH
YDIIAGSTLYPYGITEGFLPMSDTACTVIAHTADYNECSNGLCDRIKIVLCECPPG
YDGAACORASCPKANSSTPGSGGERSNTNFKVFANSAPFHGRAAEVVRDQCSCGHG
TCMTIEQLAFIDHGNSTYDLWDKDTVM"
BASE COUNT 166 a 163 c 144 g 170 t
ORIGIN
```

Query Match

Best Local Similarity 55.2%; Pred. NO. 1.4; Length 643;
Matches 80; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY	278	TGAAGTGTCTGCTGAACCGTATCGCTGGAGGACTACCGACTACGCTGCTATCATCA	337
Db	282	TCTCGGATACTGCTGGAACCGTTATCGCTAACACGCTCATGACTACATGAATGTTCCA	341
QY	338	CCGAGTGTGTGAAGTGTGTCATCACTTCTACACGACGACGCTTCACTTCAACGCTG	397
Db	342	ATAAGGATATTCGATCGCATCAATGCTTTTGGGAATGTCCTCGCTATGACGCTG	401
QY	398	GAGCTTCTACCTGTATCCGCTGTGCC 422	
Db	402	CAGCTTCCAGCGTGCATCATGTCC 426	

RESULT 10

AC099748

LOCUS

AC099748 Bos taurus clone RP42-504H16, WORKING DRAFT SEQUENCE, 5 ordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos. 1 (bases 1 to 179217) Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S. M., Benjamin, B., Blakesley, R. W., Bouffard, G. G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N. L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Idol, J. R., Karlins, E., Laric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q. L., Maduro, V. B., Margulies, E. H., Masello, C., Maskeri, B., Mastrian, S. D., McCloskey, J. C., McDowell, J., Paguirigan, C., Pearson, R., Portnoy, M. E., Prasad, A., Schueler, M. G., Stantripop, S., Thomas, J. W., Thomas, P. J., Touchman, J. W., Tsurgeon, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E. D. NISC Comparative Sequencing Initiative

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (15-JUN-2002) NIH Intramural Sequencing Center, 8717 Government Circle, Gaithersburg, MD 20877, USA On Jun 15, 2002 this sequence version replaced gi:17017552. ----- Genome Center Center: NIH Intramural Sequencing Center Center code: NISC Web site: http://www.nisc.nih.gov Contact: nisc_zoo@hgri.nih.gov

----- Project Information
Center project name: crq
Center clone name: 504H16

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 177994 bases at least Q40
Consensus quality: 178622 bases at least Q30
Consensus quality: 178774 bases at least Q20
Insert size: 136000; agarose-fp
Insert size: 178817; sum-of-contigs
Quality coverage: 12.94x in Q20 bases; agarose-fp
Quality coverage: 9.85x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 32314: contig of 32314 bp in length
32315 32414: gap of unknown length
32415 66136: contig of 33722 bp in length
66137 66236: gap of unknown length
66237 80633: contig of 14397 bp in length
80634 80733: gap of unknown length
80734 92700: contig of 11967 bp in length
92701 92800: gap of unknown length
92801 179217: contig of 86417 bp in length.

Location/Qualifiers

1 .179217
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="RP42-504H16"
/clone_lib="RP42"
1 .32314
/notes="assembly_fragment
clone_end:SP6
vector_side:left"

misc_feature

32415..66136
/note="assembly_fragment"
66237..80633
/note="assembly_fragment"
80734..92700
/note="assembly_fragment"
92801..179217
/note="assembly_fragment
clone_end:T7
vector_side:right"
164910..179217
/note="clone overlaps with GenBank Accession Number
AC13228 clone RP42-67B23 (center project name crl)"

BASE COUNT 53762 a 38501 c 36655 g 49899 t 400 others
ORIGIN

Query Match 2.9%; Score 41; DB 2; Length 179217;

```
Best Local Similarity 50.2%; Pred. No. 2.3;
Matches 101; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 843 TACCTGTCTGCTTGTCTGCTTAAACAAGACTACGAGCTAGGCTACCGTGGAGGAGC 902
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 106670 TAGCTAGCAGACCTTCGCGAGGTTACAAGCAGCGTCTGTGCTGCTGCTGCTG 106729

QY 903 TGCTAGCCCTGGTAAAGCTAGTAAACATCGCTTGCTGCTGACGAACCGCTATCGCTTCGG 962
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 106730 TGTAAAGTCGCTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 106789

QY 963 AGCTACCAACTACGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1022
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 106790 CTCCTCCCACTCCCTGGGATCTCCAGCAAGAACTGAGTGGTGGCATTTCTCTCTC 106849

QY 1023 CGACGGAAACAACTTCGAGC 1043
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 106850 CAATGCATGAAAGTGAAAGC 106870

RESULT 11
AF374552 591 bp mRNA linear PLN 27-JUL-2001
LOCUS
DEFINITION
Thalassiosira weissflogii isolate CMP 1336 clone 13 sexually
induced protein SIG 1 mRNA, partial cds.
ACCESSION
AF374552
VERSION
AF374552.1 GI:14586994
KEYWORDS
SOURCE
Thalassiosira weissflogii.
ORGANISM
Thalassiosira weissflogii
Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
Thalassiosirophyceae; Thalassiosirales; Thalassiosiraceae;
Thalassiosira.
REFERENCE
1 (bases 1 to 591)
Armbrust, E.V. and Galindo, H.M.
Rapid evolution of a sexual reproduction gene in centric diatoms of
the genus Thalassiosira
Appl. Environ. Microbiol. 67 (8), 3501-3513 (2001)
JOURNAL
MEDLINE
PUBMED
11472926
AUTHORS
Armbrust, E.V. and Galindo, H.M.
DIRECT SUBMISSION
TITLE
Submitted (27-APR-2001) Oceanography, University of Washington, Box
357940, Seattle, WA 98195-7940, USA
JOURNAL
FEATURES
source
1. 591
/organism="Thalassiosira weissflogii"
/isolate="CMP 1336"
/db_xref="taxon:67004"
/clone="13"
/country="USA: Long Island, NY"
<1..>591
/codon_start=2
/product="sexually induced protein SIG 1"
/protein_id="AAK70420.1"
/db_xref="GI:14586995"
/translation="SHDMCTCDRNFQAGDCSLRTPFGKAHVDTTPRGDLDSLITGDH
NDIILAGTSLYPYGTGEPFLMSGTAGTAVTANTAHDMYMECSNKGCLDRINGLCCLPG
YDGAACORASCPKSNYSTPGSGGERSNTNFKVFANSAFHGRAAEVYQRDQCSDHG
TCMTIEQLFLDHGNTYDLWDKDPYTC"
BASE COUNT 152 a 146 c 144 g 149 t
ORIGIN

Query Match 2.8%; Score 39.4; DB 8; Length 591;
Best Local Similarity 54.5%; Pred. No. 4.3;
Matches 79; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 278 TGAAGTGTCTGCTGGAACCGCTATCGCTGGAGGAGCTACGACTACGCTGCTATCATCA 337
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 198 TGTCCGGTACTGCTGGAACCGCTTATCGCTAAACACCGCTCATGACTATGATGAATGTCCA 257
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 338 CCGAGTGTGTGAAGTGTGCTGATCAACTTCTTACACGAGACGCTTAACTTCAACGCTG 397
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 258 ATAAGGATTATGATGATCGATCAATGGCTTTGCGAATGCTTCTTGGCTATGACGGTG 317
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 398 GAGCTTCTACCTGTACCGCTTGTC 422
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 318 CAGCTTGCCAGCGTGCATCATGTC 342
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AF374492 643 bp DNA linear PLN 27-JUL-2001
LOCUS
DEFINITION
Thalassiosira weissflogii isolate CCMP 1336 clone 9 sexually
induced protein SIG 1 gene, partial cds.
ACCESSION
AF374492
VERSION
AF374492.1 GI:14336347
KEYWORDS
SOURCE
Thalassiosira weissflogii.
ORGANISM
Thalassiosira weissflogii
Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
Thalassiosirophyceae; Thalassiosirales; Thalassiosiraceae;
Thalassiosira.
REFERENCE
1 (bases 1 to 643)
Armbrust, E.V. and Galindo, H.M.
Rapid evolution of a sexual reproduction gene in centric diatoms of
the genus Thalassiosira
Appl. Environ. Microbiol. 67 (8), 3501-3513 (2001)
JOURNAL
MEDLINE
PUBMED
11472926
AUTHORS
Armbrust, E.V. and Galindo, H.G.
DIRECT SUBMISSION
TITLE
Submitted (27-APR-2001) Oceanography, University of Washington, Box
357940, Seattle, WA 98195-7940, USA
JOURNAL
FEATURES
source
1. 643
/organism="Thalassiosira weissflogii"
/isolate="CCMP 1336"
/db_xref="taxon:67004"
/clone="9"
/country="USA: Long Island, NY"
join(<1..56,141..>643)
/product="sexually induced protein SIG 1"
join(<1..56,141..>643)
/codon_start=2
/product="sexually induced protein SIG 1"
/protein_id="AAK61134.1"
/db_xref="GI:14336348"
/translation="SHDMCTCDRNFQAGDCSLRTPFGKAHVDTTPRGDLDSLITGDH
NDIILAGTSLYPYGTGEPFLMSGTAGTAVTANTAHDMYMECSNKGCLDRINGLCCLPG
YDGAACORASCPKSNYSTPGSGGERSNTNFKVFANSAFHGRAADVQRDQCSDHG
TCMTIEQLFLDHGNTYDLWDKDVTM"
BASE COUNT 166 a 162 c 147 g 168 t
ORIGIN

Query Match 2.7%; Score 38.6; DB 8; Length 643;
Best Local Similarity 48.0%; Pred. No. 7.4;
Matches 110; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 194 CCGTGTACCCCTGTCTCTCAGAGAAGGAGCGCTGGAGCTCAGCTAACCTCTCTGTACCG 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 198 CCGTTACAATTGGCGACCAACAGACATAATTTTACGTGGTCCGACTCTTTATCATACG 257
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 254 CTAACCTGTGTGACCCAGTGTAACTGTAAGTGTCTGCTGGAACCGCTATCGCTGGAGGAG 313
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 258 GTATCACCAGGACTTCCTTTCATGTCGGATACCTGCTGAACCGTATCGCTAACACG 317
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 314 CTACCGACTACGCTGTATCATCATCAGGAGTGTGTGACCTGCGCATCACTTCTACAACG 373
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 318 CTATGACTACATGGAATGTTTCCAATAAGGATTTATCGCATCGCATCAATGGCTTTGCG 377
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 374 AGAACGCTCTACTACTCAACGCTGGAGCTTCTACCTGTACCGCTTGTC 422
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 378 AATGTCTTCTGCTATGACGGTGCAGCTTGCACGCTGCATCATGTC 426
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```
gene
CDS
EPAKGRKNASLRVSSSSHRPLTPSRVTALVLRPELI"
complement(1734..2477)
/genes="alr7004"
/genes="alr7004"
complement(1734..2477)
/genes="alr7004"
/genes="alr7004"
/notes="ORF_ID:all7004"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB78088.1"
/db_xref="GI:17135542"
/translation="MTIIEDYCSAVRSSITNDGHPPIEASGLKIQENLTIETOSLEPM
EKRSALPPLVNLKHLAKLSATSLFSPRVAYQWVDKASNLNNKIGLDAAGVYQ
SYQQLITMSQOKQAKATLNTAIDNFIKTTYSWGLFHCYEIEDFPRTNDELHAFG
MLRYHQRICTGRKVPASSLVRGSKVLCACATATKLHSFTASDLAQVDIHTWLEURSOL
OKHKRARTQYRFREDPKAYLANESRL"
complement(2527..3228)
/genes="alr7005"
complement(2527..3228)
/genes="alr7005"
/genes="alr7005"
/notes="ORF_ID:all7005"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB78089.1"
/db_xref="GI:17135543"
/translation="MWNEYNNPHRIITLNGVVELOLQIRCONKSCMRKYKAYRPQE
GSLALPQNEFGLDVATIGALRYQEHRSVPQIHTHLKLGICISKRTVTHLIDRYDEL
LSLWLKDHKRLKAIYVNOGRVILADGMQPEIGHEVLMWVRDCLSGEILAKTLTSSR
NEDLVALLLEVANTLDVPIDGVDGSDGOOSIRKAVGLALPKIAHGLCHYHYLKEAIRPI
YEADNARKELKKKS"
3495..3743
/genes="asr7006"
3495..3743
/genes="asr7006"
/genes="asr7006"
/notes="ORF_ID:asr7006"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB78089.1"
/db_xref="GI:17135543"
/translation="MWNEYNNPHRIITLNGVVELOLQIRCONKSCMRKYKAYRPQE
GSLALPQNEFGLDVATIGALRYQEHRSVPQIHTHLKLGICISKRTVTHLIDRYDEL
LSLWLKDHKRLKAIYVNOGRVILADGMQPEIGHEVLMWVRDCLSGEILAKTLTSSR
NEDLVALLLEVANTLDVPIDGVDGSDGOOSIRKAVGLALPKIAHGLCHYHYLKEAIRPI
YEADNARKELKKKS"
3495..3743
/genes="asr7006"
3495..3743
/genes="asr7006"
/genes="asr7006"
/notes="ORF_ID:asr7006"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB78090.1"
/db_xref="GI:17135544"
/translation="MRORLSKQDLAWLNKAKNSVSLRSFAVSLESDDYNNAVKGVTM
SYSGNPVEGHINRLMLKRMQYGRAKIDLLERFLITM"
3776..4126
/genes="alr7007"
3776..4126
/genes="alr7007"
/genes="alr7007"
/notes="ORF_ID:alr7007"
/hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB78091.1"
/db_xref="GI:17135545"
/translation="MSYRNQFIWQRAVOLAINCYKFTRLFPQSELYGLTSQTRRSVS
VASNIAPGGRSKPEYIQFLHALGSLRELDQTLIIAKEVDLADKDLFTPIINVEVEE
MQSLVATLNKLAG"
complement(4331..5200)
/genes="all7008"
complement(4331..5200)
/genes="all7008"
/genes="all7008"
/notes="ORF_ID:all7008"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB78092.1"
/db_xref="GI:17135546"
/translation="MRISYQYPLKPTKEALHIDKTLDMLRQYNYQLAQRDFWTEQN
RCSVDRCSLVWCHIPELRDXPNRFESQATLTQLKDRPWYKSIHSQTLQEQVPKVEIA
FDRLWLGSDNGSRKRPKRGKQYKFTYPOFKKEHFTKDKITLSKIGTKYKIVHPR
IPGDFDKTVSVTKADGYVTLSEDKTVPITKTDLDIINNIVGIDVGLIDFIVTSDG
DRAAPKFLRKAERKLKLAQRKSVSRKKSGNRKKKAIQOLGQHKHKKVADTRKDFHYKT
ANNCDFTPCNTKF"
5254..5673
gene
```

```

CDS
/genes="alr7009"
5254..5673
/genes="alr7009"
/genes="alr7009"
/notes="ORF_ID:alr7009"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB78093.1"
/db_xref="GI:17135547"
/translation="MKDDFVSKGSVSDLKVHLVLVTKYRKKAFTSEMLSRNLNVMOE
LLEKNDCKLVTFNGEEDHVHLLFQYHPDVELSKLVNNLKSVSRRKLQKQFAEHLSEFY
WKDFWSSGYFVASC GGVTLLTRKYIEAQESLTQST"
complement(5798..7252)
/genes="all7010"
complement(5798..7252)
/genes="all7010"
/genes="all7010"
/notes="ORF_ID:all7010"
similar to hemolysin export system membrane fusion protein
hlyD"
/codon_start=1
/transl_table=11
/protein_id="BAB78094.1"
/db_xref="GI:17135548"
/translation="MTDYSLSHAATPDEFLPPINRWVTMGVVLLSIFATFFFLASIL
KYRTVVKPATVPDPELRLVQAATEGKIISLAQVQVNTNOGDIILALVEDDLQTK
SOLOTNQATQLQTLQIDAQITALNKQIIAERDRINRVI SAAEAELNRNDRQDFDREI
NSSVQFRAQANLQQAELKQTSOLKSAITANFKSETSLKAAQARRDRYKPLLETG
TSQDQFOFEOVLAFEOQOILQESOTANVEQSOQAIOOKYVIAAARLQALANLNP
SDANVIAQEIATERAVSEVSLRNLQEREOLLOKLEKQKLSRDRQELNIMREI
ANTAIRASAGIIQELNLRNPAQIIRSGDIVAQISPKRSPLIICALYPASDIRKVTN
QEVQLRISSCSYIEGLKNSVSAISPDVINSGNPNPAMYEVTTIQPKLLELTAGDRQ
CAIKSGMEARADILSETIILQFLKKARLTNI"
complement(7321..8292)
/genes="all7011"
complement(7321..8292)
/genes="all7011"
/genes="all7011"
/hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB78095.1"
/db_xref="GI:17135549"
/translation="MNVLIITHSDNESISLVTOATESOGGKAFRDTDRFTEVOLD
IYYSNTEKCVLVDQDKLDNEVAVVRIAGGIPTMTDKQLRQASIQESRATIQ
GMIASIRGFHLDVPNIRRAENKQLQVARKIGLDTPTLTNNQVAKFEAACQO
DVTIKMLSSFAIYDEKGEQVFTNPVKSEDLNLEGLRCPMTFOEK TAKVLELRIT
IVGKSILTAAVNSQALDKSKRYDMRKQGVALLDAWQHTLPPQDVADKLQLMAHFGLY
GAIDVILTPDNRYVLELVNVPGEFFWLRCPCGLPISQALAKVLLSHI"
complement(8304..9284)
/genes="all7012"
complement(8304..9284)
/genes="all7012"
/hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB78096.1"
/db_xref="GI:17135550"
/translation="MPLQRDVVLLIITHSGDYFTIDRVAALSRNNVQSFRDLTDKFFPM
TVKIQAYFHQSNSHHQIEYGDITLNTQEQVAVMMRRLWQPHLSPELAQYRDACKTES
LAVDGFWDLSLRHAHWDDLOKINAENKLYOLRVAAEVLVLPPTLVNPNKAREF
FEQVNGKMITKLLKPLSYSGESSFPMYTVKKEEDLLDAETLRYCPMVFOQOIPKQO
ELRQVYVNGNLFGALDASSYEAASQDWRRAQESCTWQPELKEITQIHLQDQFQMARL
GLTFGAFDFIVTPEEYVLEINPTGEMGWMLERDLNYPISAEIADSLIQN"
complement(9429..9731)
/genes="all7013"
complement(9429..9731)
gene
CDS
Query Match 2.7%; Score 38.6; DB 1; Length 341950;
Best Local Similarity 44.6%; Pred. No. 12;
Matches 152; Conservative 0; Mismatches 189; Indels 0; Gaps 0;
QY 121 CCTGCTAACTGTGTGAACACTGTCAGAGAAGAACTTCTACTACAACAACGCTGCTGCTTTCGGG 180
```

